(12) UK Patent Application (19) GB (11) 2 137 791 A

(43) Application published 10 Oct 1984

(21) Application No 8233119

(22) Date of filing 19 Nov 1982

(71) Applicant
The Secretary of State for Defence, (United Kingdom), Whitehall, London, SW1A 2HB

(72) Inventors
John Scott Bridle, Richard Martin Chamberlain

(74) Agent and/or address for service
Michael Holt, Procurement Executive, Ministry of Defence, Patents 1A(4), Room 2014, 20th Floor, Empress State Building, Lillie Road, London, SW6 1TR

(51) INT CL[3]
G10L 1/00

(52) Domestic classification
G4R 11A 11D 11E 12F 1F 3B 3C 8F 8X 9B PE RM
U1S 2322 G4R

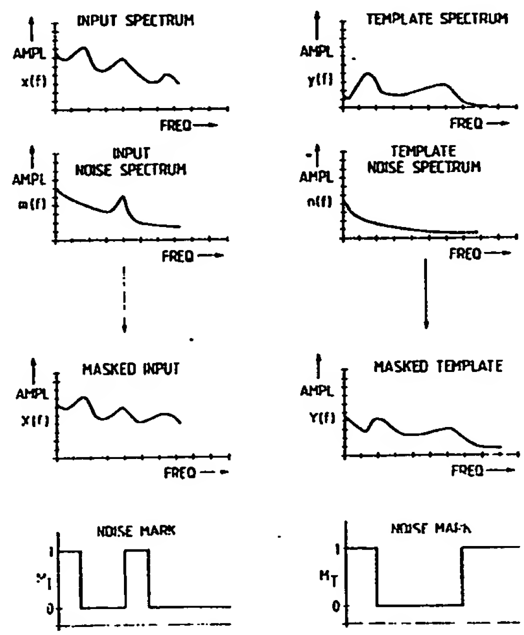
(56) Documents cited
None

(58) Field of search
G4R

(54) Noise Compensating Spectral Distance Processor

(57) A spectral distance processor for preparing an input speech spectrum and a template spectrum for comparison, as for example in pattern matching by spectral distance computation, has means for masking the input spectrum (x(f)) with an input noise spectrum estimate (m(f)), means for masking the template spectrum (y(f)) with a template noise spectrum estimate (n(f)) to give masked spectra (x(f), y(f)), and means for marking samples of each masked spectrum with a noise mark (M_I, M_T), for example 1(speech) or 0(noise), dependent upon whether the sample is estimated to be due to speech or noise. Such noise marked spectra may then be used in spectral distance pattern recognition algorithms. The noise mark may be used to adjust normalisation applied to the spectra before comparison or to recognize that a spectral distance may be due to noise by substituting a default distance for the actual distance should the greater of the masked spectrum samples be marked as noise.

Fig.3.



GB 2 137 791 A

The drawings originally filed were informal and the print here reproduced is taken from a later filed formal copy.

Fig. 1.

(PRIOR ART)

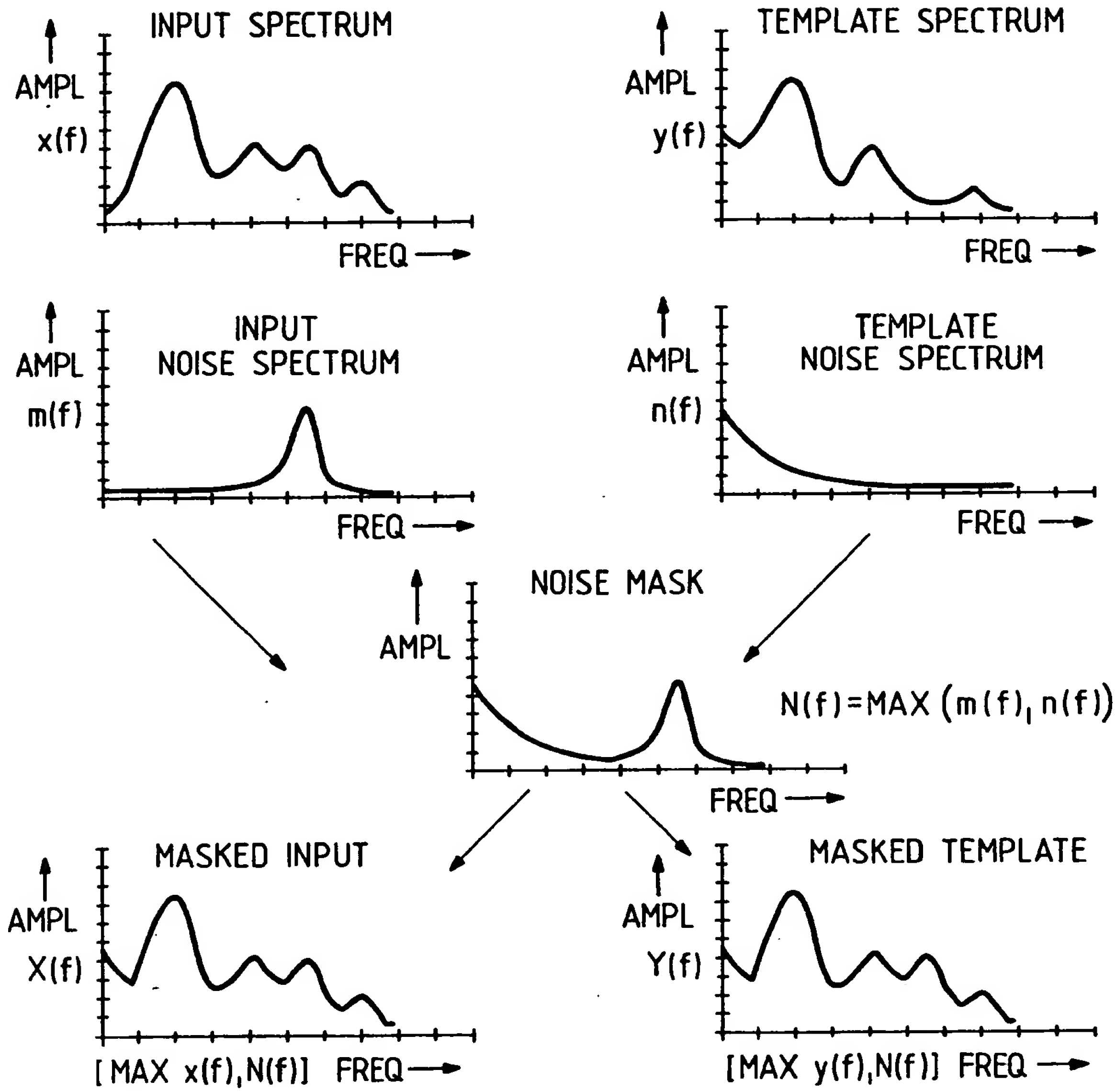


Fig. 2.

(PRIOR ART)

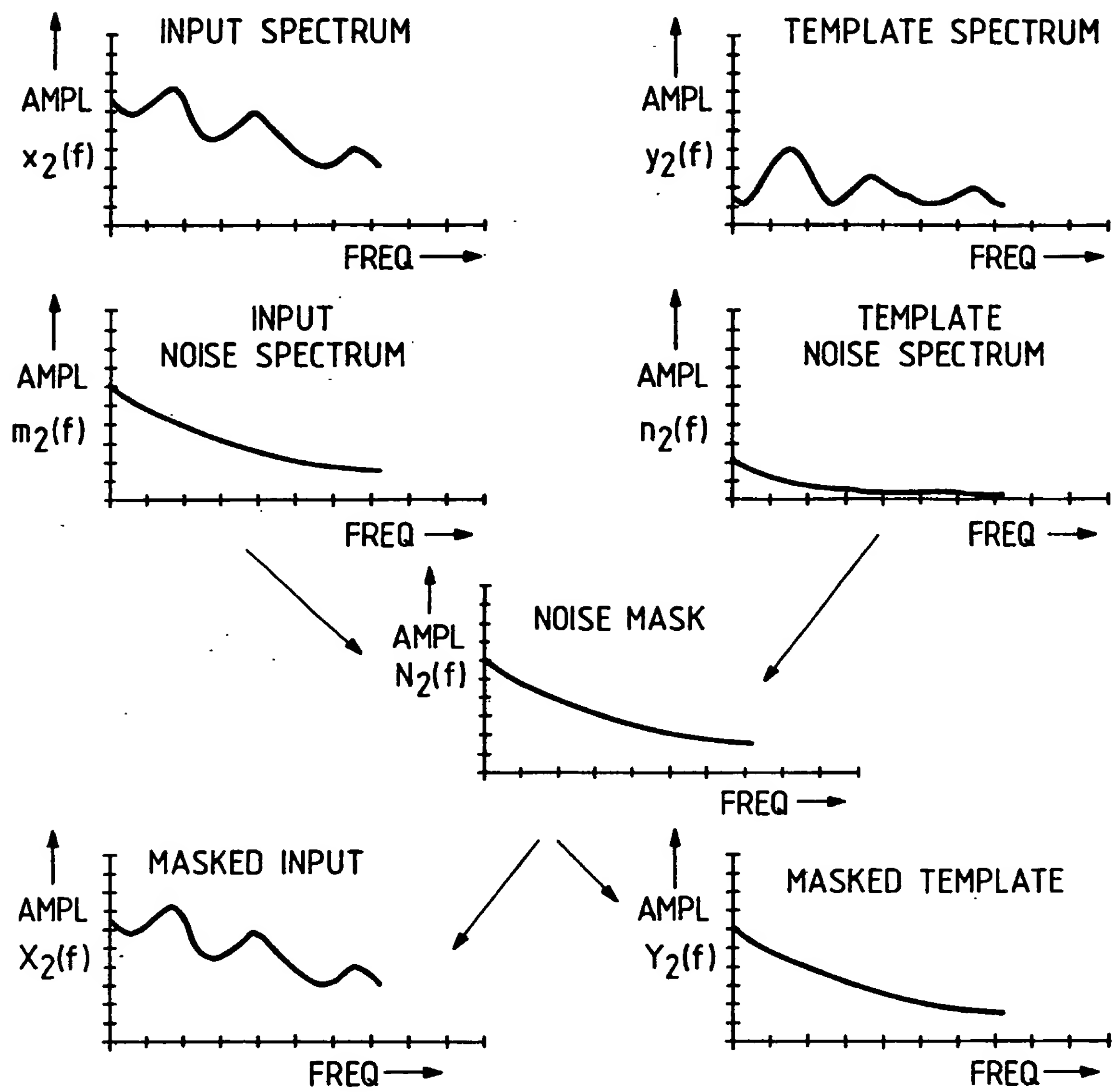


Fig.3.

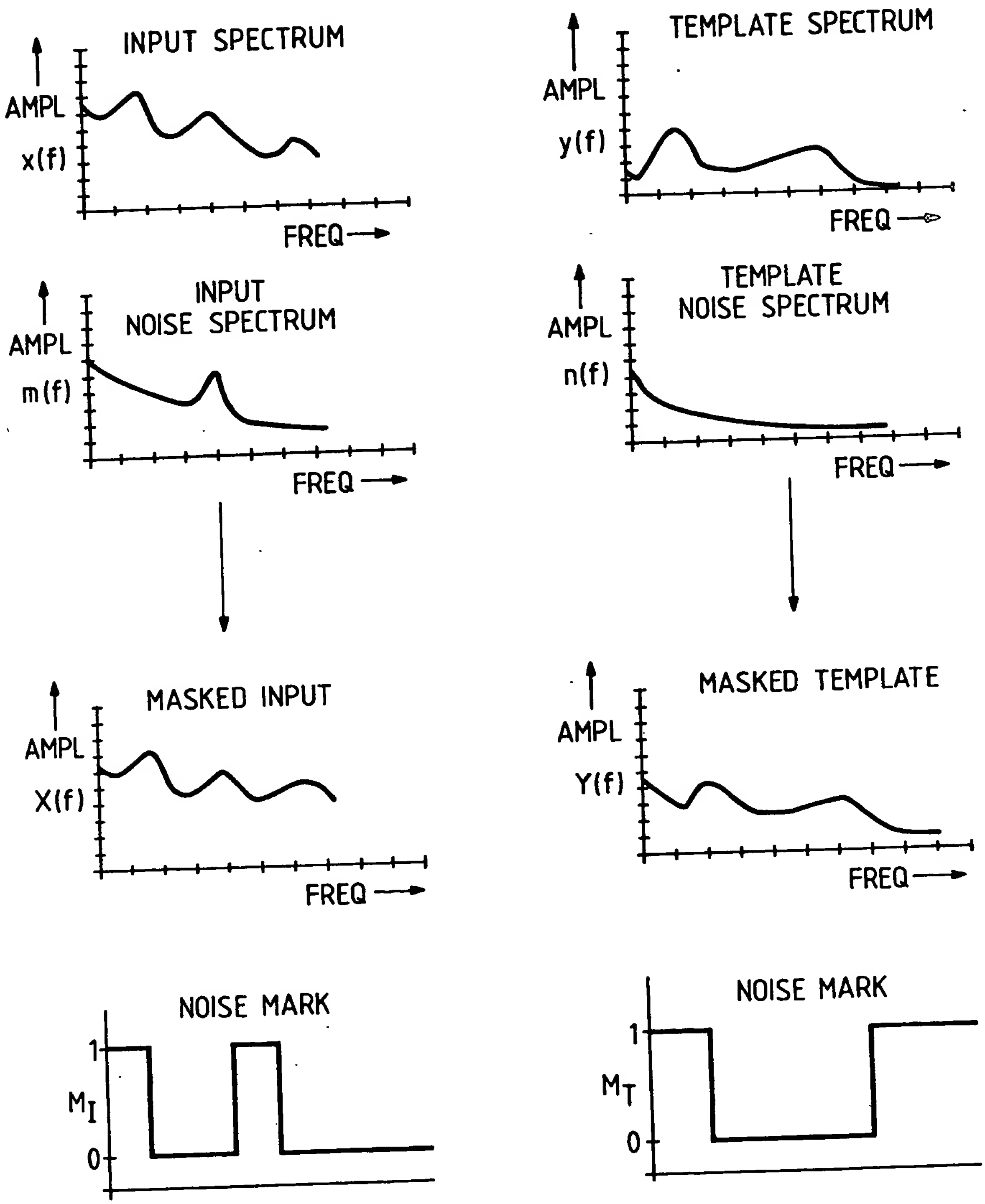
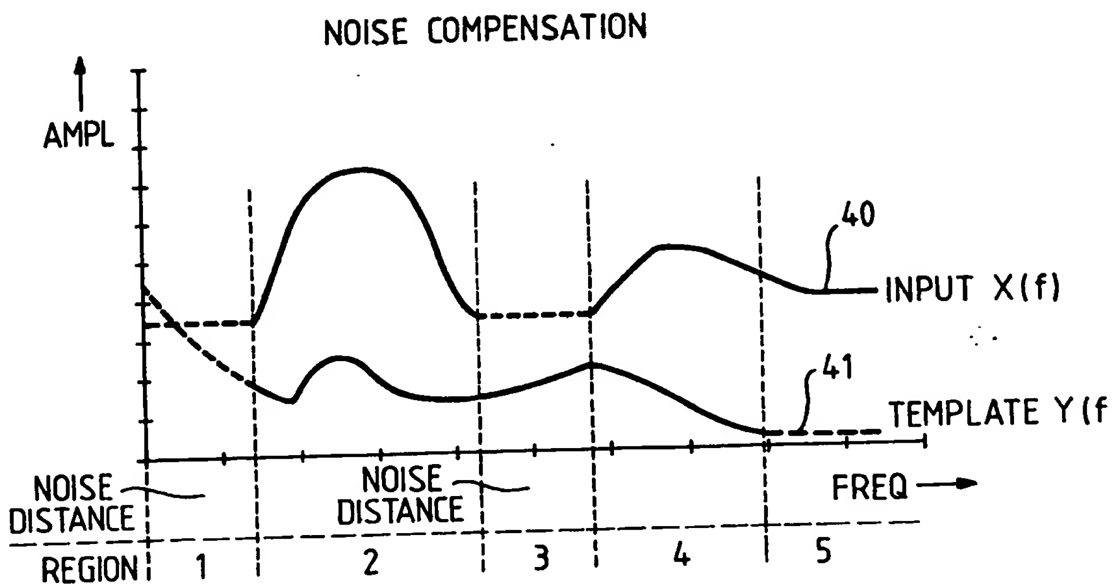

INPUT SPECTRUM
AMPL
x(f)
FREQ
TEMPLATE SPECTRUM
AMPL
y(f)
FREQ
INPUT
NOISE SPECTRUM
AMPL
m(f)
FREQ
TEMPLATE
NOISE SPECTRUM
AMPL
n(f)
FREQ
MASKED INPUT
AMPL
X(f)
FREQ
MASKED TEMPLATE
AMPL
Y(f)
FREQ
NOISE MARK
1
M_I
0
NOISE MARK
1
M_T
0

Fig. 4.


NOISE COMPENSATION
AMPL
40
INPUT X(f)
41
TEMPLATE Y(f)
NOISE DISTANCE
NOISE DISTANCE
FREQ
REGION
1
2
3
4
5

SPECIFICATION

Noise Compensating Spectral Distance Processor

This invention relates to spectral distance processors and in particular to spectral distance processors for comparing spectra taken from speech in the presence of background noise.

Speech can be represented as a sequence of spectra which are measures of power at various frequencies. In a speech recognition system spectra from unknown input words are compared with spectra from known templates or references.

An important practical problem in automatic speech recognition is dealing with interfering noise, such as background noise, non-speech sounds made by a speaker and intrucive sounds of short duration such as a door slamming. In general input and template spectra will be obtained in different noise environments to compound the problem of comparison.

In order to provide speech recognition in the presence of noise the technique of noise masking has been proposed. The basis of the technique is to mask those parts of the spectrum which are thought to be due to noise and to leave unchanged those parts of the spectrum estimated to be speech. Both input and template spectra are masked with respect to a spectrum made up of maximum values of an input noise spectrum estimate and a template noise spectrum estimate. In this way spectral distance between input and template may be calculated as though input and template speech signals were obtained in the same noise background.

Unfortunately known masking techniques have a number of draw backs. In particular the presence of a high noise level in one spectrum can be cross coupled to mask speech signals in the other. Four spectra are required in the spectral distance calculations, making any implementation extremely computation intensive and limiting the practicality of the technique for automated speech recognition.

According to the present invention a spectral distance processor for preparing an input spectrum and a template spectrum for comparison includes:

means for masking the input spectrum with respect to an input noise spectrum estimate,

means for masking the template spectrum with respect to a template noise spectrum estimate, and

means for marking samples of each masked spectrum dependent upon whether the sample is due to noise or speech.

The masked spectra may be used for spectral distance calculations in accordance with known and documented principles. Advantageously the spectra may be normalised before distance calculations are performed.

In a preferred form of the present invention, where the greater of the masked spectral samples is marked to be due to noise a default noise distance is assigned in place of the distance between the two masked spectra.

In an alternative form, each spectral sample is marked with a weighting, dependent upon the likelihood of that sample being due to signal and not noise.

A developed version of the present invention for speech recognition advantageously included in a speech recognition system.

In order that features and advantages of the present invention may be appreciated examples will now be described with reference to the accompanying diagrammatic drawings, of which:

Figure 1 represents prior art noise masking;

Figure 2 represents prior art noise masking;

Figure 3 represents noise masking in accordance with the present invention, and

Figure 4 represents noise masking in accordance with the present invention.

In the examples considered two spectra for comparison are referred to as the input and template spectra and their log power spectra denoted by $x(f)$ and $y(f)$ respectively, where f is frequency. Estimates of the spectra of the background noise in the input and template are denoted by $m(f)$ and $n(f)$ respectively. In the figures the spectra are drawn as continuous functions, but in practice we would be typically dealing with the outputs from a bank of band-pass filter analysis channels.

In order that the background to the present invention may be appreciated examples of prior art noise masking will now be considered. A detailed account has been given by D. H. Klatt in "A digital filter bank for spectral matching", (Proc Int Conf Acoustics, Speech and Signal Processing, pp 573—576, April 1976).

Figure 1 illustrates the prior art procedure for the case of two identical underlying spectra in different noise backgrounds.

From the two noise estimates, a noise spectrum mask is calculated by:

$$N(f)=\mathrm{Max}\,(m(f), n(f))$$

The input and template spectra are then masked by the composite noise spectrum to produce the modified spectra:

$$X(f)=\mathrm{Max}\,(x(f), N(f))$$

$$Y(f)=\mathrm{Max}\,(y(f), N(f))$$

The intention is to make new input and template spectra which appear to have the same noise background and so that they can be compared directly using the standard distance calculation. Figure 1 shows that the method has indeed produced two similar spectra for comparison, $X(f)$, $Y(f)$.

There are problems with the above method. A theoretical problem is that, due to effectively masking one spectrum with the noise estimate of the other spectrum, meaningful differences

between the two spectra which were apparent may be hidden. For instance, if there is high background noise in one signal, then the masking of the other signal may lessen the difference seen in the data. This can happen because the level of power in the two spectra is different, even though it is only the shape of the two spectra that we want to compare. A practical problem is that the calculation of the noise-masked distance requires four spectra, and the spectrum distance is the most computation-intensive operation in a pattern-matching speech recogniser.

In another example of the method (Figure 2), the technique fails to provide spectra suitable for comparison ($X_2(f)$, $Y_2(f)$) since a high level of noise in the input spectrum ($M_2(f)$) is coupled via noise mask $N_2(f)$ to masked template $Y_2(f)$.

In accordance with the present invention an input spectrum x(f) (Fig. 3) is masked with an estimate of input noise m(f) to give a masked input X(f) such that:

$$X(f)=\max(x(f), m(f))$$

A template spectrum y(f) is similarly masked with noise estimate n(f) to give a masked template Y(f) such that:

$$Y(f)=\max(y(f), n(f))$$

It will be appreciated that if background noise is stationary then masking will have little effect. The masking will however be useful in fluctuating or high noise level conditions. It will further be appreciated that cross-coupling of noise via the masking process cannot occur.

During the masking operations noise marks M_I and M_T are associated with the masked spectra X(f), Y(f) respectively according to whether the value arose from noise (noise mark 1) or speech (noise mark 0) and taken into account during spectral distance calculations on X(f) and Y(f).

The way in which masked spectra X(f) and Y(f) may be compared will now be described graphically with reference to Fig. 4.

The input 40 and template 41 spectra are plotted on the same axes and the parts of the spectra that are considered to be noise (noise mark 1) are drawn in dashed lines, while the solid lines represent the parts of the spectra that are thought to be speech. It will be appreciated that the noise spectra are no longer required for this distance calculation.

The usual distance function is denoted by F(X—Y) e.g. $F(X-Y)=(X-Y)^2$.

A spectral distance calculation, modified to include information about the noise may now be performed as follows:

If, at any frequency channel, the larger of X(f) and Y(f) is due to the noise, as in Regions 1 and 3 of Figure 4 then the channel distance, D, is given by

$$D=D^* \qquad \text{(a)}$$

where D^* is a default noise distance. In this case nothing can be deduced about the difference between the two spectra at this frequency channel. Instead of assigning a zero value (which denotes a perfect match) to the distance for such a channel, D is given the non-zero value D^*. In this way a perfect match can only be found between spectra that are identical after normalisations and not from a comparison of two spectra that are just noise.

If the maximum of X(f) and Y(f) is due to the signal, as in Regions 2, 4 and 5 of Figure 4, then the channel distance is given by

$$D=F(X(f), Y(f)) \qquad \text{(b)}$$

It will be realised that this equation uses all the available information from the channel because, even if the lower level is due to noise, the difference between the two signal levels must be at least that in (b). In the special case when the higher level is due to signal, the lower level is due to noise and the value of D from (b) is less than D^*, then we assign the value D^* as the distance for that channel. The distance between the two spectra can now be found by adding together the values of D from (a) or (b) for each channel.

This algorithm may be implemented simply in hardware since after the spectra have been marked for signal or noise, all that has to be stored is the decision of the marking for each channel and this only requires one bit. Thus the noise compensation is not just part of the acoustic analysis but also an integral part of the distance calculation.

Before the spectrum distance is calculated various spectral normalisations may advantageously be applied. Usually an amplitude normalisation is carried out by subtracting a proportion of the means from the two spectra. However, even the amplitude normalisation may be adversely affected by the background in that the estimates of the mean may be distorted by the noise in some channels. The most comprehensive method of applying the amplitude normalisation in the present invention is to calculate the estimate of the mean from those channels which are considered to be speech in both the template and the input. This involves a significant amount of computation and this can be reduced considerably by subtracting off a proportion of the peak channel level, which should be due to the speech.

A generalised form of the present invention will now be described in which each masked spectrum sample is marked with a weight, which is an indication of the likelihood that the channel value is due to signal rather than noise.

The weight is found by comparing the channel value with the noise estimate. Letting the weights for the input and template be denoted by WX(f) and WY(f), these weights are associated with the masked spectra and can be used in the various spectral normalisations. By making this extension to the invention the implementation now requires more than one bit for the weight for each channel. The spectrum distance calculation is then

modified so that the channel distance is weighted between the normal distance and the default noise distance:

$$D=W(f)\cdot F(X(f),Y(f))+(1-W(f))\cdot D^{*} \quad (c)$$

where W(f) is the weight of the higher signal, that is

$$W(f)=WX(f)\ldots\ldots\ldots\ldots\text{if } X(f)\geqslant Y(f),$$

or

$$W(f)=WY(f)\ldots\ldots\ldots\ldots\text{if } X(f)<Y(f)$$

Again the spectrum distance is just the sum over all the channels of all the values of D from (c).

In this way the distance calculation is now continuous as the noise level rises, since the weights adjust gradually to changes in the estimates of the noise level. However, the distance is still discontinuous when the input and template values are nearly equal and their weights are different. This can be simply solved by introducing a slight change so that, when the channel values of the input and template are nearly equal, both WX(f) and WY(f) are used in calculating D.

Though this continuous version of the method does not require a hard decision about the signal, it does require extra storage for the weights and more computation to use them. A compromise can be taken by using just a few bits for the weights, for instance two bits may be adequate. This retains the advantages of the continuous version of the method without adding too much to the storage and computation requirements.

CLAIMS (Filed on 18/11/83)

The matter for which the applicant seeks protection is:

1. A spectral distance processor for preparing an input spectrum and a template spectrum for comparison including means for masking the input spectrum with respect to an input noise spectrum estimate, means for masking the template spectrum with respect to a template noise spectrum estimate, and means for marking samples of each masked spectrum dependent upon whether the sample is due to noise or speech.

2. A spectral distance processor as claimed in claim 1 and including means for performing spectral distance calculations to compare the masked spectra.

3. A spectral distance processor as claimed in claim 2 and including means for normalising the spectra before comparison.

4. A spectral distance processor as claimed in claim 2 or claim 3 and including means for assigning a default distance in place of the calculated distance whenever the greater of the masked spectrum samples is marked as due to noise.

5. A spectral distance processor as claimed in any preceding claim and including a single bit of store for storing the noise mark associated with each sample.

6. A spectral distance processor as claimed in any of claims 1 to 4 and wherein the noise mark associated with each sample is a weighting dependent upon the likelihood of that sample being due to speech not noise.

7. A spectral distance processor as claimed in claim 6 and including storage bits for storing a noise mark weighting associated with each sample and wherein the number of storage bits is less than the number of bits required to fully specify the weighting.

8. A speech recognition system including a spectral distance processor as claimed in any preceding claim.

9. A speech recognition system as claimed in claim 8 and including a plurality of frequency restricted channels, each channel having a spectral distance processor as claimed in any of claims 1 to 7.

10. A speech recognition system as claimed in claim 9 and including means for normalising the spectra in each channel with respect to an estimate of the mean from those channels which are marked to be speech in both input and template spectra.

11. A spectral distance processor substantially as herein described with reference to Figs. 3 and 4 of the drawings.

Printed in the United Kingdom for Her Majesty's Stationery Office, Demand No. 8818935, 10/1984. Contractor's Code No. 6378. Published by the Patent Office, 25 Southampton Buildings, London, WC2A 1AY, from which copies may be obtained.